Advanced Encryption Approaches for Machine-to-Machine Communications and Big Data Processing

Miodrag Mihaljević
Mathematical Institute,
Serbian Academy of Sciences and Arts

Workshop on Mathematical Data Science

Belgrade, Serbia, 22 June 2015

Roadmap

- Introduction
- Certain Security Challenges: Minimization of the Security Overheads
- Illustrative Technical results: Towards Advanced Lightweight Security Mechanisms

1. Introduction

Our Topic, Challenges and Messages

Our Topic: IoT and Big Data

Our Topic: IoT & M2M Communications

Our Topic:

M2M Is All About Real-Time "Big" Data

© http://www.zdnet.com/big-data-all-you-need-to-know-1339335818/

http://strata.oreilly.com/2012/01/what-is-big-data.html

Challenges

- Current information-communication technologies (ICT) heavily involve machine-to-machine (M2M) communications and big data processing.
- These two issues open a number of mathematical challenges regarding data privacy and secrecy.
- A particular challenge is developing of algorithms which support reduction of the overheads implied by employed techniques for data security. Frequent employment of the security mechanisms could result into a heavy cumulative overhead regarding implementation complexity, computational complexity and power consumption.
- At the same time we face a request for a high and preferably provable security of the employed cryptographic techniques.

Main Messages

- Accordingly, the advanced cryptographic techniques should at the same time provide **high security and as small as possible overheads to the main functionality of a system**.
- This talk considers elements of an encryption approach for fulfilling the claimed goals based on **involvement elements of coding theory into traditional compact cryptographic techniques for encryption** in order to provide a secure and lightweight processing of data for the privacy and secrecy purposes.

- We point out to certain elements from **coding theory** which shows that **joint employment of pseudo-randomness generated by compact finite-state machines, randomness and dedicated coding provide a framework for developing of lightweight and provable secure encryption algorithms for data privacy and secrecy.**

2. Certain Security Challenges

Minimization of the Overheads
Implied by Security Mechanisms

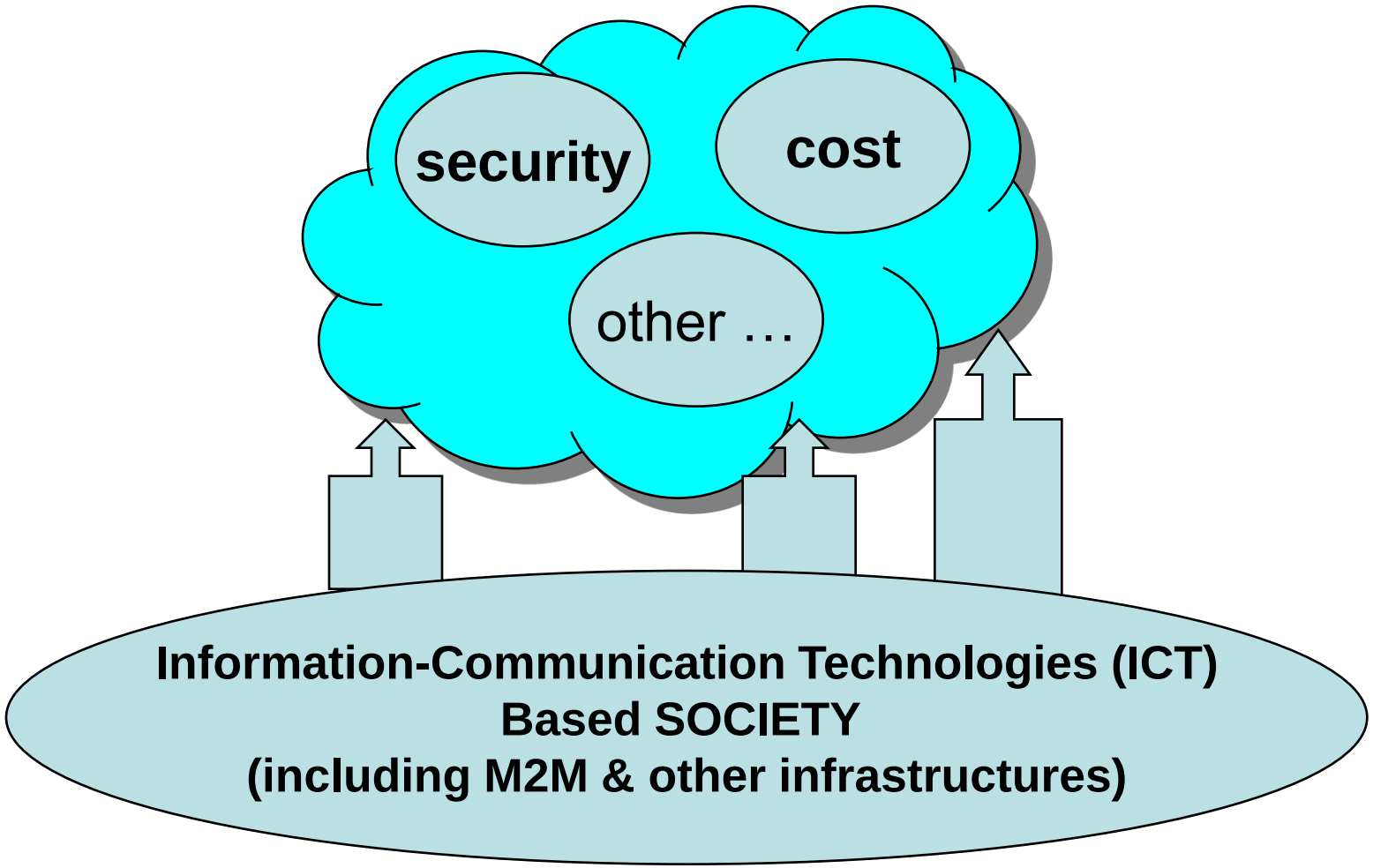
IoT Atributi

Security – The Key Issue of IoT

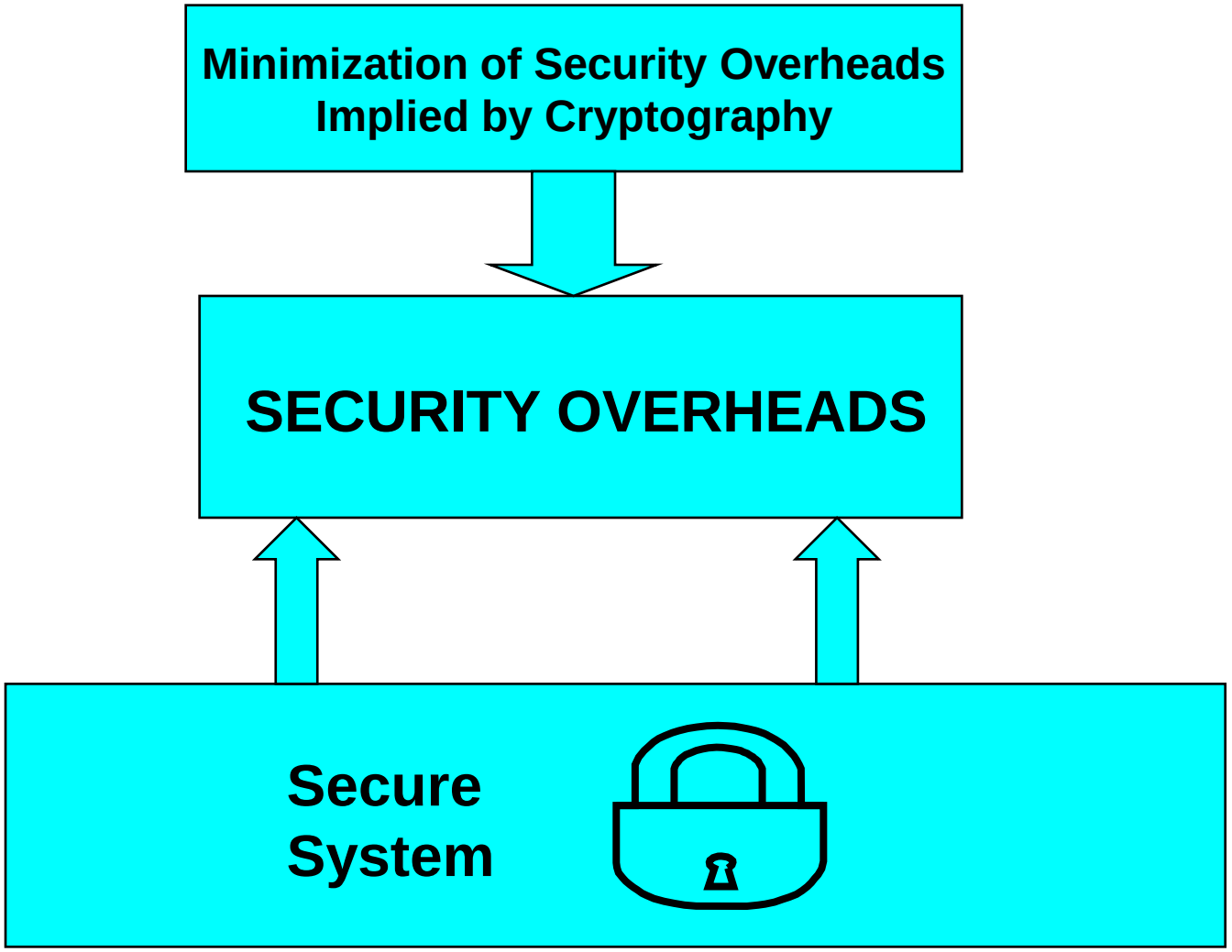
Certain Challenges of Cyber-Security

- **Security evaluation of the components and basic lightweight cryptographic algorithms**
- **Design** of advanced lightweight cryptographic algorithms which provide **provable security and low implementation overheads**

IoT Concerns



Minimization of the Security Costs



Minimization of the Overheads Implied by Security Mechanisms

- minimization of the implementation overheads
- minimization of the computational overheads (and particularly: minimization of the key management, authentication and encryption overheads)
- minimization of the communications overheads
- minimization of power-consumption

Trends & Implications

- **Extensive employment of cyber-security mechanisms**
- **Overheads to the main functionality implied by employed security mechanisms**

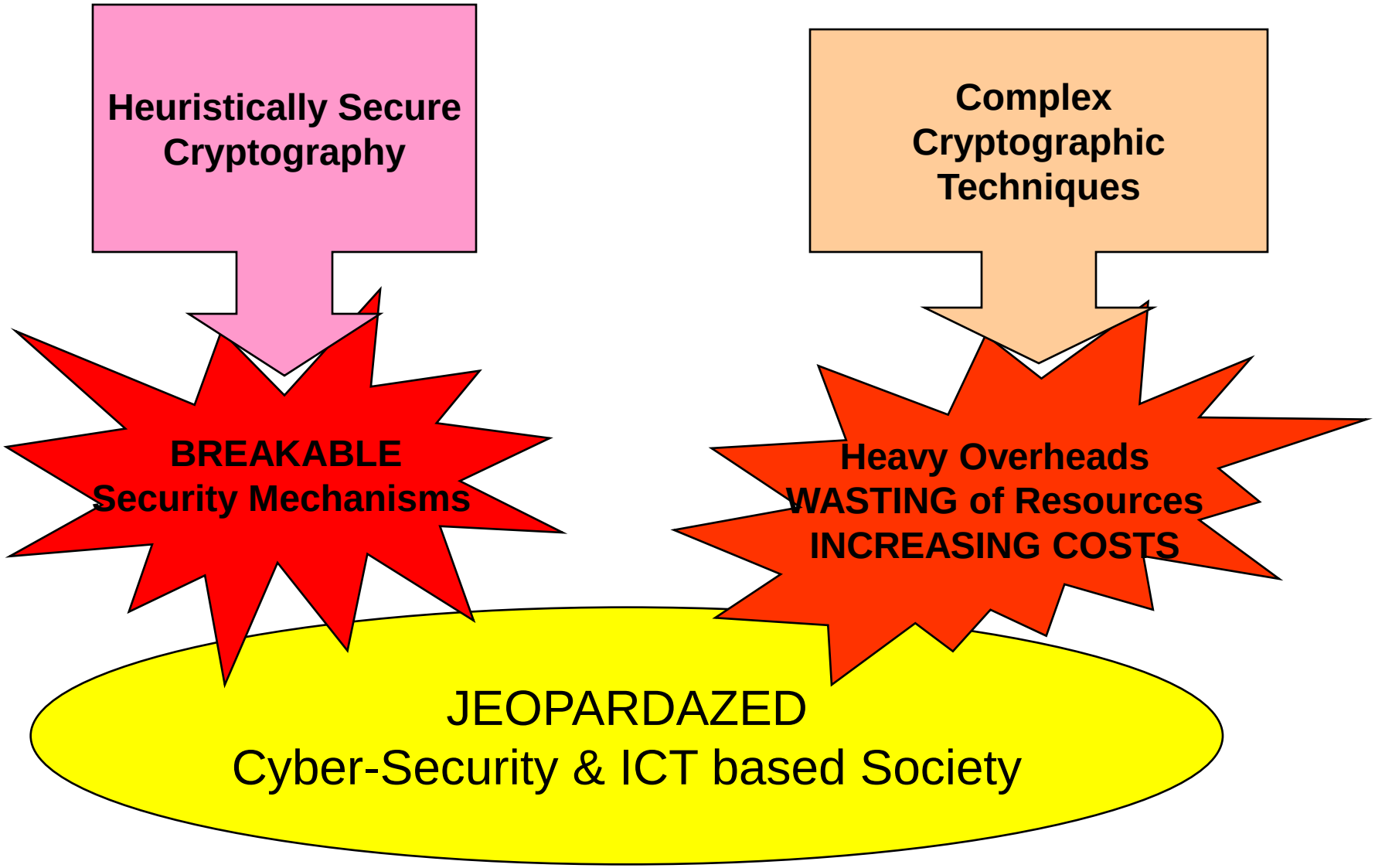
Cryptographic Challenges

Provable Security of employed security mechanisms

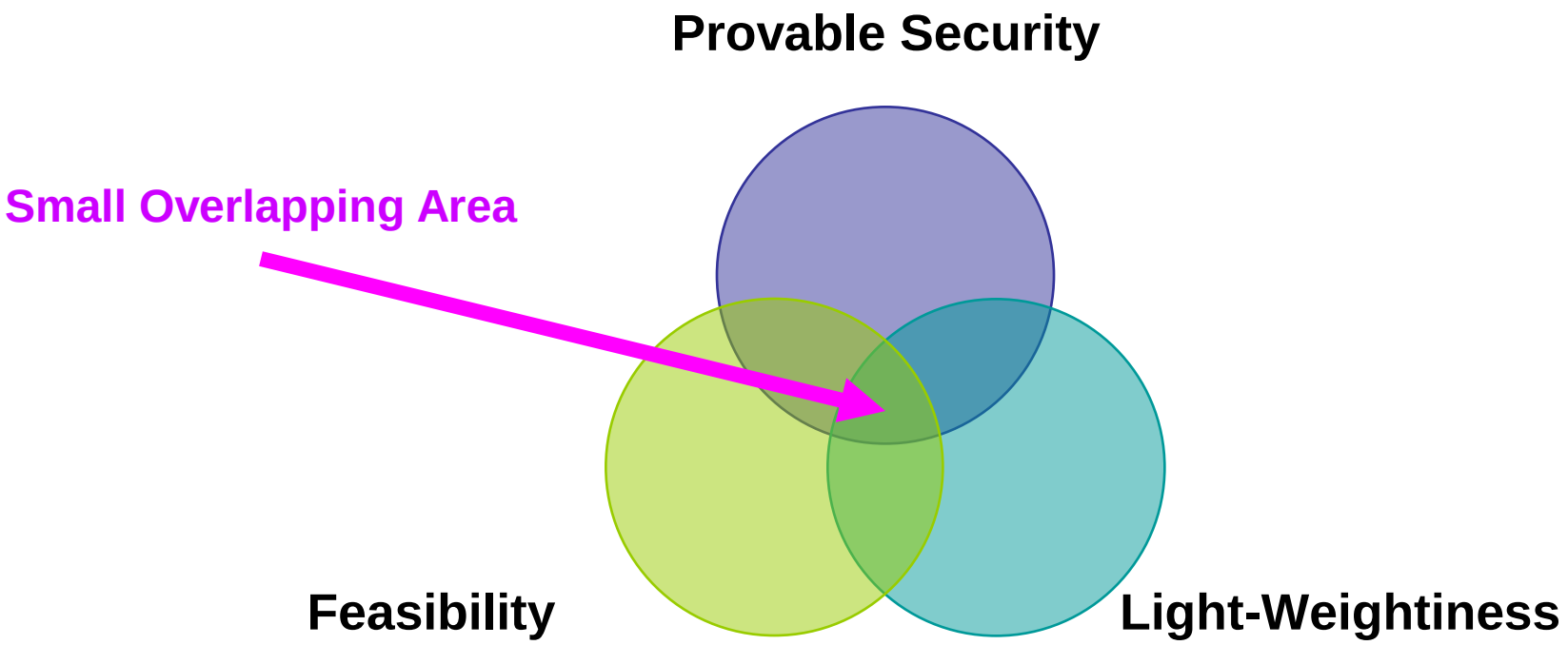
Provable Secure means that breaking is as hard as solving a hard mathematical problem

Minimization of the overheads implied by the security mechanisms

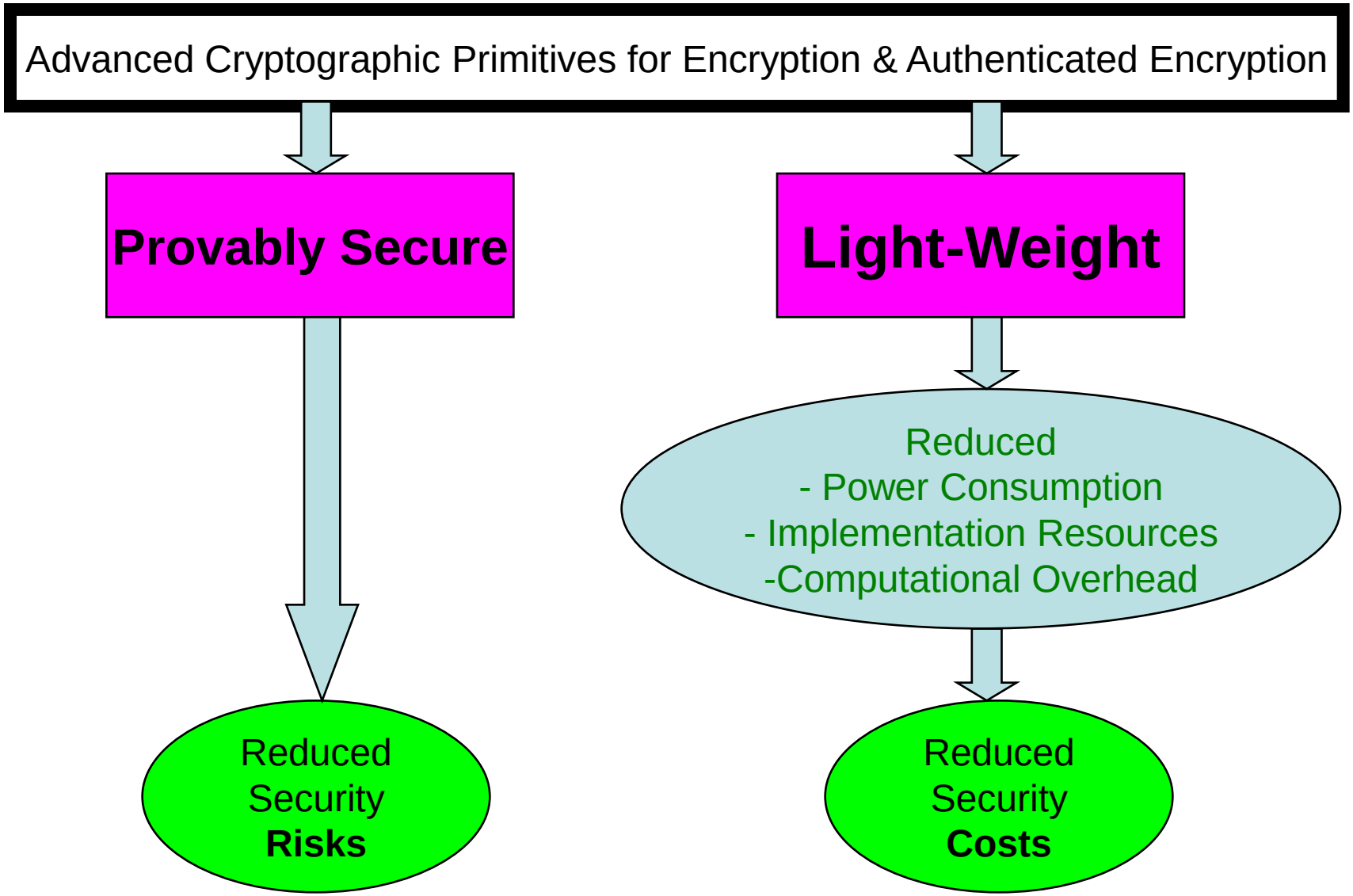
Potential Disastrous Impacts



Cryptographic Challenges regarding Cryptographic Primitives



Particular Goals of the Advanced Construction



3. Illustrative Technical Issues

References (1)

- F. Oggier and M.J. Mihaljevic, "**An Information-Theoretic Security Evaluation of a Class of Randomized Encryption Schemes**", *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, pp. 158-168, Feb. 2014.
- M.J. Mihaljevic and H. Imai, "**Security Issues of Cloud Computing and an Encryption Approach**", in *High Performance and Cloud Computing in Scientific Research and Education*, IGI Global, US, pp. 388-408, March 2014.
- M.J. Mihaljevic, S. Gangopadhyay, G. Paul and H. Imai, ``**Internal State Recovery of Keystream Generator LILI-128 Based on a Novel Weakness of the Employed Boolean Function**", *Information Processing Letters*, vol. 112, no. 21, pp. 805-810, Nov. 2012.
- M.J. Mihaljevic, S. Gangopadhyay, G. Paul and H. Imai, ``**State Recovery of Grain-v1 Employing Normality Order of the Filter Function**", *IET Information Security*, vol. 6, no. 2, pp. 55-64, June 2012.
- M.J. Mihaljevic, S. Gangopadhyay, G. Paul and H. Imai, ``**Generic Cryptographic Weakness of k-normal Boolean Functions in Certain Stream Ciphers and Cryptanalysis of Grain-128**", *Periodica Mathematica Hungarica (Selected Papers of 2011 Central European Conference on Cryptology),* vol. 65, no. 2, pp. 205-227, Dec. 2012.
- **Japan Patent JP 4863283**: M.J. Mihaljevic and H. Watanabe, **Authentication system using light-weight authentication protocol**, November 18, 2011.

References (2)

- **United States Patent US 8023649**: M.J. Mihaljevic and J. Abe, **Method and apparatus for cellular automata based generation of pseudorandom sequences with controllable period**, September 2011.
- M.J. Mihaljevic and H. Imai, ``**An approach for stream ciphers design based on joint computing over random and secret data**'', *Computing*, vol. 85, no. 1-2, pp. 153-168, June 2009.
- M.J. Mihaljevic, M. Fossorier and H. Imai, ``**Security Evaluation of Certain Broadcast Encryption Schemes Employing a Generalized Time-Memory-Data Trade-Off**'', *IEEE Communications Letters*, vol. 11, no. 12, pp. 988-990, Dec. 2007.
- M. Fossorier, M.J. Mihaljevic and H. Imai, ``**Modeling Block Encoding Approaches for Fast Correlation Attack**'', *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4728-4737, Dec. 2007.
- M.J. Mihaljevic, ``**Generic framework for secure Yuen 2000 quantum-encryption employing the wire-tap channel approach**'', *Physical Review A*, vol. 75, no. 5, pp. 052334-1-5, May 2007.

Advanced Approach Against Existing Ones

- Practical light-weight cryptographic primitives based on **randomized algorithms**.
- Randomization provides framework for **provable security**.

- Traditional light-weight cryptographic primitives are based on **deterministic** algorithms.
- Security is based on the **heuristics** and the “expected” **computational complexity** arguments.

Security Evaluation / Cryptanalysis

- **Security evaluation / cryptanalysis of particular lightweight symmetric-key cryptographic techniques** for application in critical infrastructures
- Developing of **generic techniques for security evaluation / cryptanalysis** of certain classes of symmetric-key encryption and authentication schemes

Illustrative Technical Issues on Advanced Lightweight Encryption

3.1 Illustrative Technical Result I

The LPN Problem Based Security

Definition: LPN Search Problem. Let s be a random binary string of length l. We consider the Bernoulli distribution $\mathcal{B}_\theta$ with parameter $\theta \in (0, 1/2)$. Let $\mathcal{Q}_{s,\theta}$ be the following distribution:

$$\{(a, < s, a > \oplus e) | \leftarrow \{0,1\}^l, e \leftarrow \mathcal{B}_\theta\} .$$

For an adversary $\mathcal{A}$ trying to discover the random string s, we define its advantage as

$$\mathsf{Adv}_{\mathsf{LPN}_{\theta,\mathcal{A}}}(l) = \Pr[\mathcal{A}^{\mathcal{Q}_{s,\theta}} = s | s \leftarrow \{0,1\}^l] .$$

The LPN_θ problem with parameter θ is hard if the advantage of adversaries $\mathcal{A}$ that make a polynomial number of oracle queries is negligible.

Definition: LPNDP - LPN Decisional (Distingushing) Problem. Let s, a be binary strings of length l. Let further $\mathcal{Q}_{s,\theta}$ be as in Definition of the LPN search problem. Let $\mathcal{A}$ be a adversary. The distinguishing-advantage of $\mathcal{A}$ between $\mathcal{Q}_{s,\theta}$ and the uniform distribution $\mathcal{U}_{l+1}$ is defined as

$$\mathsf{Adv}_{\mathsf{LPNDP}_{\theta,\mathcal{A}}}(l) = \mathsf{Pr}[\mathcal{A}^{\mathcal{Q}_{s,\theta}} = s | s \leftarrow \{0,1\}^l] - \mathsf{Pr}[\mathcal{A}^{\mathcal{U}_{l+1}} = 1] \; .$$

The LPNDP_θ with parameter θ is hard if the advantage of adversaries $\mathcal{A}$ is negligible.

It has been shown that the distinguishing-problem is as hard as the search-problem with similar parameters.

Underlying Problem of the LPN

(on hardness of solving an over-defined system of N probabilistic equations with K unknown variables)

Complexity of the problem strongly depends on the parameters K, N and the corruption noise

noisy variables

OVERDEFINED SYSTEM

linear-f1(x_1, x_2, …, x_K) = z_1

linear-f2(x_1, x_2, …, x_K) = z_2

⋮

linear-fN(x_1, x_2, …, x_K) = z_N

Left-hand side of the system Is "available" only after a binary symmetric channel with known probability of errors

K << N

Definition 1. Let $\mathbf{S}$ be a secret $k \times n$ binary matrix, and $\mathbf{P}_0$ be a secret $n \times n$ secret permutation matrix. Let $\mathbf{a}_t$ be a k-dimensional random vector which is publicly available, $t = 1, 2, ...$ Finally, let $\mathbf{P}_t = f(\mathbf{a_t}, \mathbf{P}_{t-1})$, where $f(\cdot)$ is a suitably selected function. For $t = 1, 2, ...,$ encryption of $\mathbf{x}_t$ into $\mathbf{z}_t$ is

$$\mathbf{z}_t = (C(\mathbf{x}_t)||\vec{\rho_t})\mathbf{P}_t \oplus \mathbf{a}_t \cdot \mathbf{S} \oplus \vec{\nu_t} \ ,$$

and accordingly, decryption of $\mathbf{z}_t$ into $\mathbf{x}_t$ is as follows:

$$\mathbf{x}_t = C^{-1}(tcat_m((\mathbf{z}_t \oplus \mathbf{a}_t \cdot \mathbf{S})\mathbf{P}_t^{-1}) \ .$$

Theorem. Assume there is an adversary $\mathcal{A}$, running in time T , and attacking the stream cipher specified by Definition 1 with parameters (ℓ, m, k, n, η) in the sense of IND-P1-C0 with advantage δ by making at most q queries to the encryption oracle. Then there is an algorithm $\mathcal{M}$ making $O(q)$ oracle queries, running in time $O(T)$, and such that

$$\left|\mathsf{Pr}\left[\mathrm{s} \leftarrow \{0,1\}^k \ : \ \mathcal{M}^{\Pi_{s,\eta}}(1^k) = 1\right] - \mathsf{Pr}\left[\mathcal{M}^{\mathcal{U}_{k+1}}(1^k) = 1\right]\right| \geq \frac{\delta}{n} \, .$$

What the theorem tell us?

- The theorem tell us that if we can “break” the encryption scheme in the sense of the distinguishability, than we can solve a related decisional LPN problem.

- The theorem tell us that if we can “break” the encryption scheme in the sense of the distinguishability, than we can solve a related decisional LPN problem

3.2 Illustrative Technical Result II

Information-Theoretic Security

Significance of Information-Theoretic Security Evaluation

- Information-theoretic security evaluation tell us about a **residual uncertainty (the equivocation)** of unique recovering the secret elements even with unlimited computational power.
- Note that even if, after all, the residual uncertainty does not exist, i.e. the equivocation is zero, **the underlying computational problem could be a very complex one**.

Algebraic Representation of Encryption - Linear Codes are Employed

$$C_H(\mathbf{a}||\mathbf{u}) = [\mathbf{a}||\mathbf{u}]\mathbf{G}_H, \tag{1}$$

and $\mathbf{G}_H$ is an $m \times m$ matrix, and thus

$$\begin{aligned} C_{ECC}(C_H(\mathbf{a}||\mathbf{u})) &= C_{ECC}([\mathbf{a}||\mathbf{u}]\mathbf{G}_H) \\ &= [\mathbf{a}||\mathbf{u}]\mathbf{G}_H\mathbf{G}_{ECC} \\ &= [\mathbf{a}||\mathbf{u}]\mathbf{G} \end{aligned} \tag{2}$$

where $\mathbf{G}_{ECC}$ is an $m \times n$ binary generator matrix corresponding to $C_{ECC}(\cdot)$, and $\mathbf{G} = \mathbf{G}_H\mathbf{G}_{ECC}$ is an $m \times n$ binary matrix summarizing the two successive encodings at the encryption side, implying that

$$\mathbf{z} = [\mathbf{a}||\mathbf{u}]\mathbf{G} \oplus \mathbf{x} \oplus \mathbf{v}\ . \tag{3}$$

Lemma 1. When $H(\mathbf{X}^n) = n \geq H(\mathbf{U}^{m-\ell}) + H(\mathbf{V}^n)$, the equivocation of the keystream output knowing the plaintext and the received signal fulfills the following tight lower bound:

$$H(\mathbf{X}^n|\mathbf{A}^l, \mathbf{Z}^n) \geq H(\mathbf{U}^{m-l}) + H(\mathbf{V}^n) - \delta(C_{ECC}) \ ,$$

where for a given positive $\epsilon \rightarrow 0$

$$\delta(C_{ECC}) = H(\epsilon) + \epsilon \log(2^{m-l} - 1) \rightarrow 0 \ .$$

Theorem 1. Let $H^*(\mathbf{K}|\mathbf{A}^{\tau l}, \mathbf{Z}^{\tau n})$ denote $H(\cdot)$ in the special case when $m = l$, i.e. when the pure random bits are not involved. When $\mathsf{Pr}(V_i^{(j)} = 0) = 1 - \mathsf{Pr}(V_i^{(j)} = 1) = 1 - p > 1/2$, $i = 1, 2, ..., n$, $j = 1, 2, ..., \tau$, we have

$$H(\mathbf{K}|\mathbf{A}^{\tau l}, \mathbf{Z}^{\tau n}) - H^*(\mathbf{K}|\mathbf{A}^{\tau l}, \mathbf{Z}^{\tau n}) = H(\mathbf{U}^{\tau(m-l)}|\mathbf{A}^{\tau l}, \mathbf{Z}^{\tau n}) - H(\mathbf{U}^{\tau(m-l)}|\mathbf{A}^{\tau l}, \mathbf{K}, \mathbf{Z}^{\tau n}) ,$$

and there exists a threshold τ_{thres} such that

$$H(\mathbf{K}|\mathbf{A}^{\tau l}, \mathbf{Z}^{\tau n}) \begin{cases} > 0 & \text{for} \quad \tau < \tau_{thres} \\ \to 0 & \text{for} \quad \tau \geq \tau_{thres} , \end{cases}$$

assuming that $n \leq |\mathbf{K}| \ll \tau_{thres} n$, and $\tau_{thres} np \ll \tau_{thres} n/2$.

Implications of Lemma 1 and Theorem 1

- Lemma 1 justifies that the **considered encryption model provides a certain level of the posterior uncertainty about its secret elements (i.e., positive equivocation)** assuming that the sample for cryptanalysis is (very) short.
- Theorem 1 shows that **this positive equivocation cannot be preserved, when the sample for cryptanalysis increases**, and that it tends to zero assuming a large enough sample implying that the uncertainty about the secret key employed in the keystream generator tends to zero, i.e., it can be correctly recovered by **(potentially hard)** computational efforts.

3.3 Illustrative Technical Result III

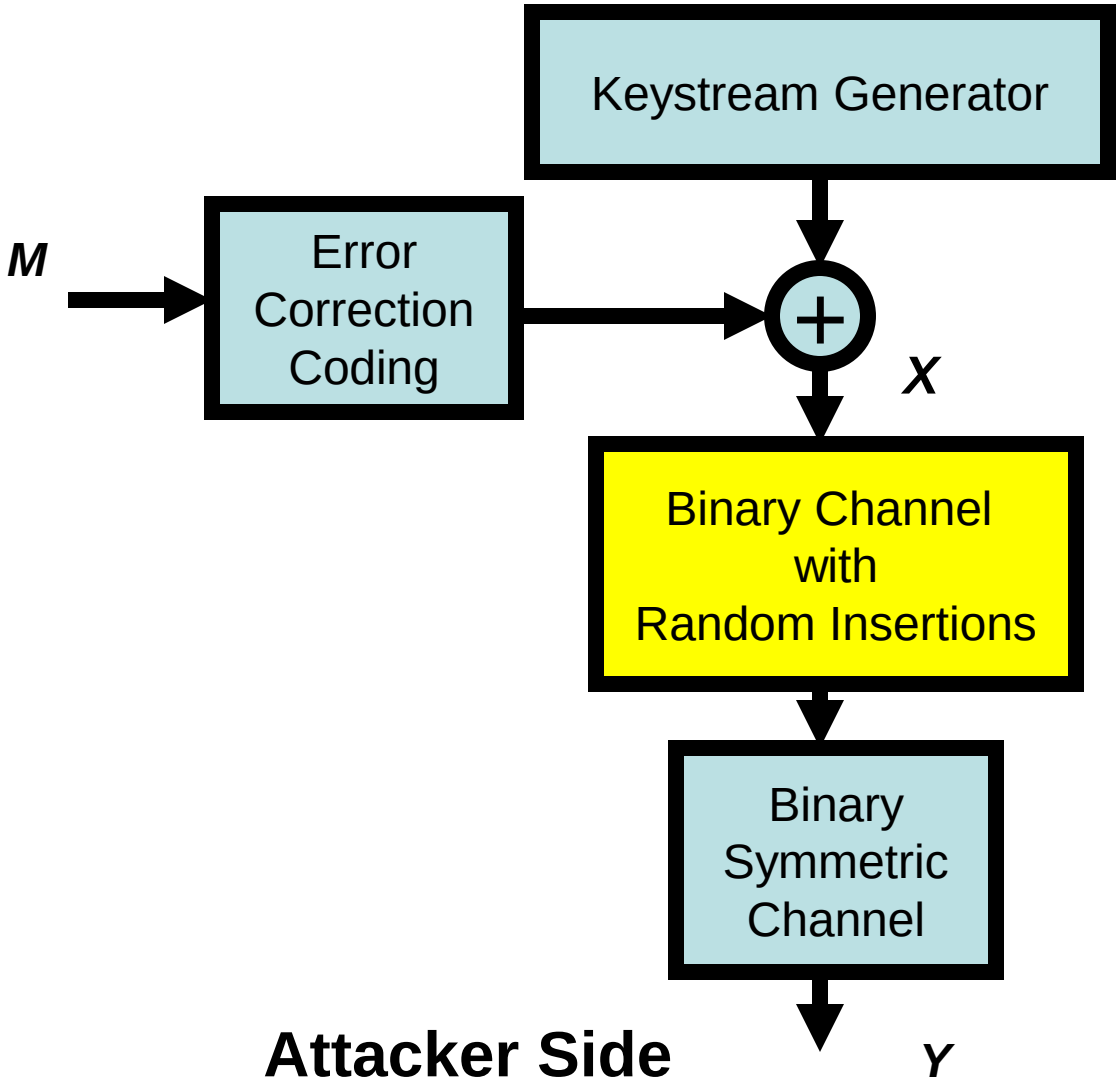
Effect of Channels with
Synchronization Errors

Underlying Idea for Security Enhancement

- **Make the channel toward the Attacker "more noisy".**

- Instead that an Attacker faces Binary Symmetric Channel, enforce that the Attacker faces a **binary channel with insertions and complementations**.

Framework Model for the Security Enhancement: the Model an Attacker's have to Face



Definition 1: The Adversarial Indistinguishability Experiment. consists of the following steps:

1. The adversary $\mathcal{A}$ chooses a pair of messages $(\mathbf{m}_0; \mathbf{m}_1)$ of the same length n, and passes them on to the encryption system for encrypting.

2. A bit $b \in \{0,1\}$ is chosen uniformly at random, and only one the two messages $(\mathbf{m}_0; \mathbf{m}_1)$, precisely $\mathbf{m}_b$, is encrypted into ciphertext $\text{Enc}(\mathbf{m}_b)$ and returned to $\mathcal{A}$;

3. Upon observing $\text{Enc}(\mathbf{m}_b)$, and without knowledge of b, the adversary $\mathcal{A}$ outputs a bit b_0;

4. The experiment output is defined to be 1 if $b_0 = b$, and 0 otherwise; if the experiment output is 1, denoted shortly as the event $(\mathcal{A} \to 1)$, we say that $\mathcal{A}$ has succeeded.

Definition 2. An encryption scheme provides indistinguishable encryptions in the presence of an eavesdropper, if for all probabilistic polynomial-time adversaries $\mathcal{A}$

$$\mathsf{Pr}[\mathcal{A} \to 1 | \mathsf{Enc}(\mathbf{m}_b)] \leq \frac{1}{2} + \epsilon \,,$$

where $\epsilon = \mathsf{negl}(n)$ is a negligibly small function.

Theorem: Let the encrypted mapping of M^n into X^n be such that $\frac{1}{2}+\epsilon$ equals the advantage of the adversary $\mathcal{A}$ (specified by Definition 2) to win the indistinguishability game (specified by Definition 1), and let the mutual information $\mathcal{I}_{iud}(X;Y)$ be known. Under these assumptions, for large n,

$$\Pr[\mathcal{A} \to 1 | Y^{(n)} = \mathbf{y}] = \frac{1}{2} + \epsilon \cdot \delta, \quad \text{where}$$

$$\delta < \mathcal{I}_{iud}(X;Y) + \frac{\log_2\left[\frac{8\pi \mathrm{e} \cdot i \cdot n}{(1-i)^2}\right]}{2n} + O\left(n^{-2}\right).$$

Implication of the Theorem

- When an attacker faces a channel with synchronization errors its problem of cryptanalysis becomes more complex.
- Deliberate insertion of random bits into the cipher-text significantly enhances the cryptographic security.

Instead Conclusions

Bart Preneel, President of IACR

(International Association for Cryptologic Research):

Challenges for Crypto

- **security for 50-100 years**
- authenticated encryption of Terabit/s networks
- **ultra-low footprint / power / energy**
- secure software and hardware implementations
- performance
- **low-cost security**

Final Note

We need Synergy of Complementary Topics for Obtaining the Additional Value

(literally to achieve: 1 + 1 => 3)

Thank You Very Much for the Attention,

and

QUESTIONS Please!